Molecular representations in AI-driven drug discovery: a review and practical guide

Abstract

The technological advances of the past century, marked by the computer revolution and the advent of high-throughput screening technologies in drug discovery, opened the path to the computational analysis and visualization of bioactive molecules. For this purpose, it became necessary to represent molecules in a syntax that would be readable by computers and understandable by scientists of various fields. A large number of chemical representations have been developed over the years, their numerosity being due to the fast development of computers and the complexity of producing a representation that encompasses all structural and chemical characteristics. We present here some of the most popular electronic molecular and macromolecular representations used in drug discovery, many of which are based on graph representations. Furthermore, we describe applications of these representations in AI-driven drug discovery. Our aim is to provide a brief guide on structural representations that are essential to the practice of AI in drug discovery. This review serves as a guide for researchers who have little experience with the handling of chemical representations and plan to work on applications at the interface of these fields.